APPLIED AND ENVIRONMENTAL MICROBIOLOGY, Sept. 1997, p. 3614–3621 Vol. 63, No. 9
0099-2240/97/$04.00+0

Diverse Uncultivated Bacterial Groups from Soils of the Arid Southwestern United States That Are Present in Many Geographic Regions

CHERYL R. KUSKE,* SUSAN M. BARNS, AND JOSEPH D. BUSCH†

Environmental Molecular Biology Group, Life Sciences Division, Los Alamos National Laboratory, Los Alamos, New Mexico 87545

Received 7 April 1997/Accepted 23 June 1997

We have performed a phylogenetic survey of microbial species present in two soils from northern Arizona. Microbial DNA was purified directly from soil samples and subjected to PCR amplification with primers specific for bacterial 16S rRNA gene sequences (rDNAs). Clone libraries from the two soils were constructed, and 60 clone inserts were partially sequenced. Phylogenetic analysis of these sequences revealed extensive diversity. Most of the analyzed sequences (64%) fell into five novel clusters having no known cultured members. Extensive analysis of 10 nearly full-length rDNAs from clones representative of the novel groups indicated that four of the five groups probably cluster into a large "supergroup" which is as distinct from currently recognized bacterial divisions as the latter are from each other. From this we postulate the existence of a major bacterial lineage, previously known only from a single cultured representative, whose diversity and ecology we are only beginning to explore. Analysis of our data and that from other rDNA sequence-based studies of soils from different geographic regions shows considerable overlap of sequence types. Taken together, these groups encompass most of the novel rDNA sequences recovered in each comparable analysis reported to date, despite large differences in soil types and geographic sources. Our results indicate that members of these new groups comprise a phylogenetically diverse, geographically widespread, and perhaps numerically important component of the soil microbiota.

Soil microbial communities are among the most complex, diverse, and important assemblages of organisms in the biosphere, yet little is known about the species that comprise them, due to the limitations of culture-based studies. One approach to defining the components of a microbial community in an environment is phylogenetic analysis of the small-subunit rRNA gene sequences (rDNAs) present in the crude pool of DNA isolated from that environment. This method allows identification of microbial species without prior cultivation, and studies using this approach in many different environments have identified new microorganisms that may be abundant or physiologically significant (2–4, 18, 20, 28, 32, 33, 38, 41). Such studies have confirmed observations from microscopic counts that the number of bacterial species we can readily culture is only a fraction of the diversity present (for reviews, see references 1, 34, and 40).

Six published rDNA sequence-based analyses of soil bacterial communities have revealed extensive diversity and many examples of novel sequences that are only distantly related to those known from cultivated species. Using PCR primers designed to amplify rDNA sequences from streptomycetes, Liesack and Stackebrandt (20) and Stackebrandt et al. (33) analyzed 113 cloned rDNAs from organisms from an acidic Australian forest soil. Surprisingly, few of these sequences affiliated with those of *Streptomyces* species; most commonly found were sequences from *Proteobacteria*, *Actinomycetes*, and planctomycetes and members of novel lineages. A more limited study of 17 partial rDNA sequences obtained from a Japanese soybean field soil by Ueda et al. (38) identified members of the gram-positive, green sulfur, and proteobacterial divisions, as well as one sequence affiliated with the domain *Archaea*. However, nine of the sequences showed no close relationship with any previously reported bacterial groups. The analysis by Borneman et al. (4) of 124 rDNA clones from a Wisconsin clover pasture soil found that they were dominated by sequences from the proteobacterial, *Cytophaga-Flexibacter-Bacteroides*, and gram-positive groups, while 39% of clones analyzed were not affiliated with any previously identified bacterial divisions. Seven clones from a grassland soil in eastern Washington State were analyzed by Lee et al. (18), and three were found to be affiliated with the *Proteobacteria*, *Flexibacter*, and planctomycete groups. Three other clones were grouped with one of the novel groups of Liesack and Stackebrandt (20), while one constitutes a unique bacterial lineage. Finally, a targeted analysis of gram-positive rDNA sequences obtained from the microbial community of a German peat bog (30) identified members of three novel lineages within the actinomycete line of descent which appear to be widespread in soils.

Many of these authors did not include sequences from other soil studies in their analyses, so potential relationships between the novel groups found in these soils have not been fully explored. Several studies have used short (~200-nucleotide [nt]) partial rDNA sequences in analyses of soil bacterial communities. While this approach can survey diversity among the sequences from a particular study site, sequences obtained from different portions of the rDNA in the various studies cannot readily be used to compare populations between environments. In addition, it is difficult to place partial sequences accurately in phylogenetic trees, especially if the sequences lack close relatives in the database. In the present study, we

* Corresponding author. Mailing address: M888, Life Sciences Division, Los Alamos National Laboratory, Los Alamos, NM 87545. Phone: (505) 665-4800. Fax: (505) 665-6894. E-mail: kuske@lanl.gov.
† Present address: Dept. of Biology, Northern Arizona University, Flagstaff, AZ 86011.

have generated nearly full-length sequences for representatives of the novel groups we have detected. We have used these sequences to infer the positions of the novel lineages in the bacterial phylogenetic tree and to compare our results with those of other soil studies.

We are studying the soil microbial communities associated with pinyon-juniper woodlands that have colonized a volcanic cinder field at Sunset Crater, Arizona, and a nearby sandy loam soil. The microbial communities in these natural areas are virtually uncharacterized, although the pinyon-juniper woodland is a major biome which extends across much of the arid southwestern United States. We have employed 16S rDNA-based molecular phylogenetic techniques to characterize a portion of the bacterial diversity in these two soils. Although meaningful comparison of species diversity between these two soil environments will require more extensive sampling and analysis, the results so far have revealed a remarkable variety of rDNA sequences, most only distantly related to those of cultivated species. Analysis of these sequences together with those obtained in other studies of soil communities reveals substantial similarities in phylogenetic types across the different soils. Most of the sequences cluster into several large, phylogenetically diverse groups that are distinct from previously recognized bacterial divisions.

MATERIALS AND METHODS

Field sites and soil conditions. The two field sites used in this study are pinyon-juniper woodlands that have similar geographic locations, elevations (approximately 7,000 ft), and general weather patterns but differ markedly in soil type. The cinder site is on the eastern edge of Sunset Crater National Monument, Arizona. The comparison site is located 12 miles due south of the Sunset Crater site, in the Coconino National Forest near the town of Cosnino. Pinyon trees used in the study were mature trees and were matched for age (160 years). At Sunset Crater the interspaces (open areas between the widely spaced trees) are largely barren. At Cosnino the interspaces are sparsely covered with grass and forb species.

The soil type of the Cosnino site is a light sandy loam (13). In comparison, the Sunset Crater site consists primarily of black, coarse-textured, well-drained cinders (5). Thus, the cinder environment generates a much drier, hotter environment for establishment of microorganisms and plants. The pHs of the two soils (7.13 to 7.57) and the levels of nitrate and ammonium nitrogen are similar. However, available organic matter, phosphorus, potassium, calcium, and magnesium are significantly lower at the cinder site. In addition, the cinder soils are more sodic than the surrounding sandy loam (16a).

Soil samples. Soil samples were collected in April 1994 to correspond with one of the two seasonal peaks of available moisture in this region. Samples were collected from the pinyon rhizosphere and interspace areas of both sites. Rhizosphere samples were collected from active feeder roots in the drip line of a tree at a depth of 10 to 15 cm, where feeder roots are most abundant. Pinyon roots were gently removed from the surrounding soil, and the soil clinging to the roots within 1 cm of the root surface was gently removed and sieved through a 2-mm mesh screen. A single composite sample was collected from each tree by pooling 10 50-cm^3 subsamples collected at evenly spaced locations around the drip line. Interspace samples were collected at the same depth, and each interspace sample represented 10 50-cm^3 subsamples collected across an approximately 1,000-ft^2 area surrounding the trees. Soil samples were mixed well, immediately placed on ice for transport, and frozen at −70°C after 24 h.

Extraction of nucleic acids from soil and cinders. Nucleic acids were extracted from two 10-g aliquots of each of the four soil samples by procedures modified from those described by Ogram et al. (25) and Tsai and Olson (37). Twenty milliliters of TENS buffer (50 mM Tris [pH 8.0], 20 mM disodium EDTA, 100 mM NaCl, 1% [wt/vol] sodium dodecyl sulfate) was added to each soil sample, and the samples were vortexed to mix thoroughly and incubated in a 70°C water bath for 1 h. Samples were mixed well at 15-min intervals during the incubation. Three grams of each of three sizes of glass beads (710 to 1,180 μm, 425 to 600 μm, and 106 μm; Sigma, St. Louis, Mo.) was added to each sample. Samples were vortexed for four 1-min intervals and then shaken vigorously back and forth for three 1-min intervals. Samples were stored on ice between the mixing intervals. Samples were centrifuged at 6,000 × *g* for 10 min at 4°C, and the supernatant was collected. The soil-bead pellet was washed once with 10 ml of TEN buffer (TENS buffer without sodium dodecyl sulfate) and centrifuged again. The wash supernatant was pooled with the original supernatant. The soil pellet was suspended in 15 ml of TEN buffer and exposed to three sets of thermal shocks by immersion of the tubes in liquid nitrogen for 2 min followed by rapid thawing in a boiling water bath. After centrifugation at 6,000 × *g*, the supernatant was collected and the soil pellet was washed as described above. Nucleic acids were precipitated from the solution (31). Precipitated samples were suspended in TE (10 mM Tris [pH 8.0], 1 mM EDTA). Protein was removed from the samples by extraction with Tris-saturated phenol followed by extraction with chloroform-isoamyl alcohol (24:1, wt/vol). The aqueous phase was collected and ethanol precipitated. DNA extracted from the soil was purified from humic acid contaminants by using Sephadex G-200 spin columns equilibrated in TE (8). The clear column eluate containing DNA was precipitated as described above and suspended in TE buffer. The DNA yield was estimated by using ethidium bromide-stained agarose gels with lambda DNA as a calibration standard.

PCR amplification of small-subunit rRNA genes. Small-subunit rRNA genes were amplified from the soil DNA by using primers conserved among all known bacteria. The primers used to initiate PCR were forward primer pA (5′-AGAG TTTGATCCTGGCTCAG-3′; *Escherichia coli* bases 8 to 27) (7) and reverse primer PC5B (5′-TACCTTGTTACGACTT-3′; *E. coli* bases 1507 to 1492) (42, 43). Test reactions were conducted with each sample to determine the optimum DNA and $MgCl_2$ concentrations for maximum yield of the 1.5-kb product. Amplification reaction mixtures used for library generation contained 10 mM Tris-HCl (pH 8.3), 50 mM KCl, 0.001% (wt/vol) gelatin, 3 mM $MgCl_2$, 200 μM each deoxynucleoside triphosphate, 100 pmol of each primer, 2.5 U of *Taq* polymerase (AmpliTaq; Perkin-Elmer, Foster City, Calif.), and 1 ng of soil DNA template in a final reaction volume of 100 μl. PCR was conducted with a Perkin-Elmer 9600 thermal cycler as follows: 2 min of denaturation at 94°C, followed by 30 cycles of 30 s at 50°C (annealing), 60 s at 72°C (extension), and 30 s at 94°C (denaturation), with a final 5-min 72°C extension step after cycling was complete. Ten replicate reactions were conducted for each of the four soil DNAs and then pooled to neutralize the effects of bias in any single thermal cycling reaction.

A 1.5-kb fragment was purified from contaminating products and from primers by electrophoretically separating the fragments on 1% agarose gels (SeaKem GTG; FMC BioProducts, Rockland, Maine) in 1× Tris-acetate-EDTA buffer. The desired fragment was visualized by staining with ethidium bromide and excised from the gel, and DNA was purified by electroelution of the gel slice in Tris-acetate-EDTA with an Elutrap chamber (Schleicher and Schuell, Keene, N.H.). DNA eluted from the gel was concentrated by using an anion-exchange Elutip-D column (Schleicher and Schuell), followed by ethanol precipitation.

Small-subunit rRNA gene libraries. Clone libraries of small-subunit rRNA gene copies from each of the four soil samples were generated. The 1.5-kb PCR product was ligated into the pGEM-T plasmid vector (Promega, Madison, Wis.) by using T4 DNA ligase and overnight incubation at 16°C. Recombinant plasmids were transformed into competent *E. coli* DH10β Electromax cells (Gibco BRL, Gaithersburg, Md.), and colonies containing plasmids with inserts were identified by blue-white color selection on agar plates (31). Individual white colonies were subjected to PCR of insert sequences to ensure that an insert of the correct size was present in each clone (10). Two hundred clones of the correct size for each of the four samples were stored as glycerol stocks at −70°C. Each clone was designated C0 (Cosnino sandy loam interspace), C1 (Cosnino sandy loam rhizosphere), S0 (Sunset Crater cinder interspace), or S1 (Sunset Crater cinder rhizosphere), followed by the clone number (1 to 200).

Determination of nucleotide sequences. Recombinant plasmids having 1.5-kb inserts were purified from overnight cultures by using plasmid miniprep columns (Qiagen, Inc., Chatsworth, Calif.). One microgram of purified plasmid DNA was used as the template in cycle sequencing reactions with fluorescent dye-labeled terminators (ABI PRISM Dye Terminator Cycle Sequencing Kit; Perkin-Elmer). Primers used for sequencing were pA (described above), P3MOD (5′-ATTAGA TACCCTDGTAGTCC-3′; *E. coli* bases 787 to 806) (42, 43), the reverse complement of P3MOD (EC910/931-RC [5′-CTCAAAGGAATTGACGGGGGC-3′; *E. coli* bases 931 to 910]), and T7 and SP6 primers corresponding to vector DNA flanking the cloned insert (Promega). Unincorporated dyes were removed from the cycle sequencing reactions by using Centri-Sep spin columns (Princeton Separations, Adelphia, N.J.). Electrophoresis of sequencing reaction mixtures was performed with 4.0% polyacrylamide gels on a 373A Stretch DNA Sequencer (Applied Biosystems, Inc., Foster City, Calif.).

The 5′ region of the 16S rDNA gene (corresponding to *E. coli* nt 62 through approximately 760) was sequenced from 15 clones from each of the four libraries, using primers pA and EC910/931-RC. Full-length insert sequences (1,463 to 1,521 bp) were obtained from 10 16S rDNA clones representative of phylogenetically novel groups of bacteria. Both strands of each DNA template were sequenced to generate a consensus sequence, and in some cases a third sequencing pass was conducted to resolve ambiguities.

Phylogenetic analysis of sequence data. Each sequence was submitted to the CHECK_CHIMERA program of the Ribosomal Database Project (RDP) (22) to detect the presence of possible chimeric artifacts (16). Questionable sequences were further investigated by examination of secondary structure and phylogenetic analyses of different portions of the sequences. Sequences were initially analyzed by using BLAST (National Center for Biotechnology Information) and SIMILARITY_RANK (RDP) (22) searches to determine the closest available database sequences. For those 18 sequences having a database relative showing a similarity value (S_{ab}) of >0.5, the highest-scoring database sequence was obtained in aligned form from the RDP and placed into an alignment with sequences of the present study. In addition, sequences representative of those major bacterial divisions not already included in the data set and novel sequences

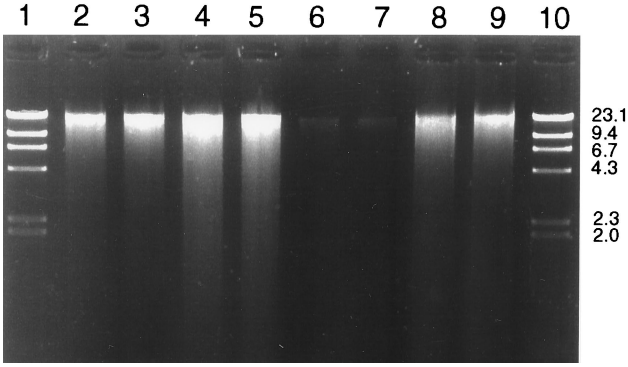

FIG. 1. DNA isolated from rhizosphere and interspace soil samples. Lanes 2 to 9, duplicate DNA samples from Cosnino sandy loam interspace (lanes 2 and 3), sandy loam pinyon rhizosphere (lanes 4 and 5), Sunset Crater cinder field interspace (lanes 6 and 7), and cinder field pinyon rhizosphere (8 and 9). Lanes 1 and 10, molecular weight markers (lambda DNA digested with *Hin*dIII) (in thousands).

from other soil studies were obtained. Alignment of the final data set was accomplished by using Clustal W (36) and manually using the GDE multiple sequence editor (22), guided by primary- and secondary-structure considerations. Sequence regions which could not be aligned with confidence were excluded from subsequent analyses. Sequence similarity calculations were performed by using PAUP (version 3.1.1; D.L. Swofford, Smithsonian Institution), based on the maximum alignable region common to all sequences (corresponding to *E. coli* nt 101 to 731; approximately 570 nt per sequence.)

Neighbor-joining distance matrix analysis of sequences was performed by using the DNADIST and NEIGHBOR programs of the PHYLIP package (version 3.5; distributed by J. Felsenstein, University of Washington, Seattle) with the Jin and Nei correction for site-to-site rate variation (15). Maximum-likelihood analyses were performed by using fastDNAml (version 1.1; distributed by RDP) (27) with empirical base frequencies and a transition/transversion ratio ($T = 1.1$) optimized by comparing likelihoods under $T = 1.0$ to 2.0. Nonbootstrap trees of full-length sequences (see Fig. 3) were found by repeated tree building with random sequence input orders and optimized by global rearrangement of branches. Trees of partial sequences (see Fig. 2) were built by random sequence addition but without branch rearrangements. Tree building was repeated until the highest-likelihood tree was found twice. In some analyses, site-specific nucleotide substitution rates were determined by using an auxiliary program, DNARATES (22), and incorporated in maximum-likelihood analysis to correct for site-to-site variations within the analysis. Maximum-parsimony analyses were performed by using PAUP, with trees found by 100 replicates of random-sequence addition heuristic searches and branch swapping on the shortest trees. Bootstrap analyses by all methods were performed on 100 replicate data sets, with random-order sequence additions. Analyses utilized Macintosh PowerPC 7600/132, SunSPARC 5, and Sun ULTRA 1 computers.

Nucleotide sequence accession numbers. The sequence data have been deposited in the NCBI database under accession numbers AF013514 to AF013568. In addition, Fig. 2 and 3, together with the alignments used to generate them, are available at http://www-ls.lanl.gov/lsdiv/LS7/SunsetCrater/SC.html.

RESULTS

DNA extraction and PCR amplification from soil samples. DNA extracted from the four soil samples was predominantly of high molecular weight (Fig. 1) and was suitable for use as a PCR template, with a minimal chance of introducing chimeric fragments due to sheared DNA. The amounts of DNA extracted from the four soil samples differed greatly. Yields of DNA from the Cosnino sandy loam site were about twice those of DNA from the Sunset Crater rhizosphere samples and about 10 times those of DNA from the Sunset Crater interspace samples. The relative DNA yields from the four samples correlated generally with total microbial counts (heterotrophic bacteria and fungi) obtained from these same samples by plating assays (16a).

rDNA clone sequencing and phylogenetic analyses. Sixty clones (15 from each soil sample) were partially sequenced (650 to 720 bp) and analyzed. Two clones were identified as putative chimeric artifacts (16, 21) based on results from CHECK_CHIMERA (17) and secondary-structure analyses (analysis not shown). Two other clones contained unusual insertion sequences of considerable length and unknown origin, and analysis of these clones is continuing. These four clones were excluded from subsequent analyses. A maximum-likelihood phylogenetic tree from analysis of sequences of the remaining 56 clones together with those of their closest relatives in the database, clones from other soil studies, and representatives of major bacterial divisions is given in Fig. 2. All sequences are of bacterial origin, as was anticipated due to the specificity of the pA forward primer. The analysis also showed that none of the recovered sequences are identical to each other (minimum sequence dissimilarity = 0.4%), and none are identical to sequences in the databases. It is clear from the preliminary analysis of clones from the four libraries that the bacterial diversity, even in the barren cinder field, is too high for critical comparisons between the different environments by this sequence-based approach.

Members of previously recognized bacterial groups. The Sunset Crater-Cosnino sequences fell into two general categories. Twenty of the 56 sequences were affiliated with previously recognized bacterial groups (26, 45) (Table 1 and Fig. 2). Sixteen of these had closest relatives among cultured taxa and clustered primarily with three divisions containing organisms commonly associated with soil: *Proteobacteria*, gram-positive organisms, and flexibacteria. Four sequences grouped within known divisions but had closest relatives among soil taxa

TABLE 1. Phylogenetic affinities of rDNA sequences obtained in this study

Phylogenetic group	Clone	Closest identified relative[a]	% Sequence identity
β-*Proteobacteria*	C005	*Rubrivivax gelatinosus*	96.3
	C107	*R. gelatinosus*	91.2
γ-*Proteobacteria*	S104	*Legionella brunensis*	91.7
α-*Proteobacteria*	S105	MC77	93.7
	C116	*Phyllobacterium myrsinacearum*	93.0
	S109	*Bradyrhizobium japonicum*	98.4
	S002	*B. japonicum*	96.8
High-G+C gram-positive bacteria	S025	*Arthrobacter globiformis*	99.6
	S007[b]	*Streptomyces ambofasciens*	86.0
	C111[b]	*S. ambofasciens*	84.0
Low-G+C gram-positive bacteria	S014	*Bacillus macroides*	91.2
Flexibacteria	S114	*Flexibacter sancti*	92.6
	C125	*F. sancti*	92.8
	S030[b]	*F. sancti*	80.2
	C113	*Flexibacter flexilis*	83.7
Planctomycetes	S019	MC98	90.2
	S101	MC98	89.6
Uncertain affiliation	C006[b]	MC77	80.0
	S017[b]	*Leptonema illini*	75.1
	C114[b]	*L. illini*	72.4

[a] Closest relative ($S_{ab} > 0.5$) identified by the SIMILARITY_RANK analysis of RDP (22).

[b] No close relative ($S_{ab} > 0.5$) identified; percent identity was determined for the closest relative in Fig. 1.

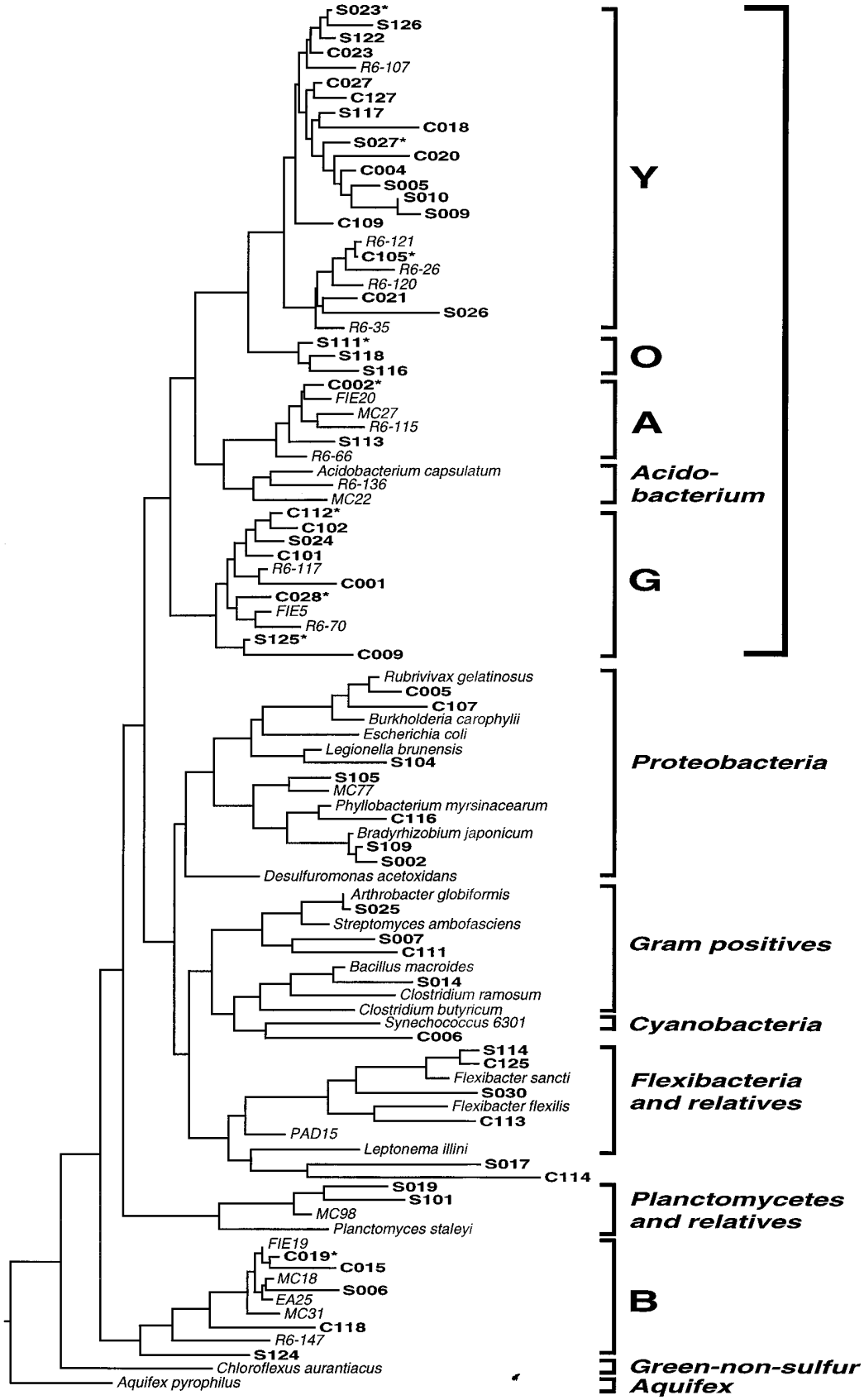

FIG. 2. Phylogenetic tree of bacterial rDNA clones from Arizona soil communities. Sequences from soil organisms (S, Sunset Crater; C, Cosnino) were analyzed with the most closely related sequences obtained from the database, as well as other representatives of major bacterial groups. Sequences of clones obtained from other analyses of soils are designated MC (20, 33), FIE (38), EA (18), and R6 (4). The tree was inferred by maximum-likelihood analysis as described in the text and was rooted by using two archaeal sequences (*Thermoproteus tenax* and *Thermococcus celer* [not shown]). Previously defined bacterial groups are named at the right based on RDP designations (22), while groups known only from environmental rDNA sequences are designated Y, O, G, A, and B. Full-length (ca. 1,500-nt) sequences were determined for clones designated with asterisks and were used in the analysis shown in Fig. 3.

known only from rDNA sequence analyses. Clone S105 showed an affiliation with Australian soil clone MC77 (designated cluster IV), which has a specific but rather distant relationship to the α-1 subdivision of *Proteobacteria* (Fig. 2) (33). Clone C006 also had the highest sequence identity to MC77, but its correct placement in the tree in Fig. 2 is unknown. The sequences of clones S019 and S101 were found to group within the family *Planctomycetaceae* (19). These two sequences form a clade rooted within soil cluster II identified by Liesack and Stackebrandt (20), exhibiting highest sequence similarity with clone MC98 of that study. Within the region sequenced, these two clones exhibit all but one of the diagnostic nucleotide signatures previously described for the planctomycetes (analysis not shown) and common to cluster II sequences (20). The relatively long branches of the S019 and S101 sequences (Fig. 2) indicate an elevated sequence mutation rate, as observed for other members of the planctomycetes group (19).

Members of novel groups. Sequences of the second category, constituting most of the analyzed clones (36 of 56, or 64%), group into five distinct clades of novel descent. These groups were originally given color codes (yellow, orange, green, and blue) to facilitate clone and data management. They are designated in Fig. 2 as Y, O, G, A, and B. Analyses of these five groups using different taxon compositions and/or analytical methods showed variation only in branching orders within groups; sequences did not move between groups. This was also true for trees constructed from data sets of full sequences, as well as those using only readily alignable positions. Similarities between these sequences and those of the most closely related cultivated organisms range from 73.0 to 82.8%, with sequence identity averaging only 79.2%. This level of similarity is comparable to that between the presently recognized bacterial divisions (i.e., average similarity between proteobacterial and gram-positive sequences in this analysis = 78.3%), and thus these groups appear to constitute novel bacterial divisions.

Analysis of novel sequences from other soils. The Sunset Crater and Cosnino sequences were analyzed with those from novel lineages in five comparable rDNA analyses of soils (4, 18, 20, 33, 38) in an effort to identify other members of the Y, O, G, A, and B groups. Many instances of overlap between the studies were identified, and most (60 to 75%) of the novel sequences identified from other studies fell into the new groups. The results of this analysis are summarized in Table 2, and representative sequences from the other soil studies are included in the tree in Fig. 2. In some instances this analysis was complicated by the short sequence length (~200 nt) in the other studies, and thus not all members of these groups may have been identified.

Of the eight clone sequences designated soil cluster VI recovered from Australian forest soil (33), four fall within group A and four group with *Acidobacterium capsulatum* (30). Twenty-two clones identified by sequencing or hybridization as belonging to soil cluster III cluster within the B group. Thus, of 42 Australian soil clones identified as belonging to "novel lines of descent" (33), 71% can be accounted for by membership in the A, B, and *Acidobacterium* groups shown in Fig. 2 and 3. Three of four clones (75%) from a natural grassland soil from Washington State (18) grouped with the sequence of Australian soil clone MC18 (18) and branched within the B group of the present analysis. Six of 10 sequences (60%) from a Japanese soybean field (38) affiliated with the A, B, and G groups. Thirty-nine sequences from an analysis of a Wisconsin clover pasture soil (4) were placed in unidentified clades A to H. At least 29 of these sequences (74%) cluster within the Y, G, A, *Acidobacterium*, and B groups.

TABLE 2. Characteristics of novel Sunset Crater and Cosnino sequence groups and affiliations with novel soil rDNA sequences from different studies

Group	No. of clones[a]	% Identity within group[a]	No. of clones from other analyses[b]			
			WI[c]	AU	JA	WA
Y	18	80.0–98.1	10	0	0	0
O	3	93.7–96.70	0	0	0	0
A	2	91.9	4	4	1	0
Acidobacterium	0		1	4	0	0
G	8	82.4–96.3	5	0	4	0
B	4 (5)[d]	83.1–94.6	6	22	1	3

[a] Sequences from Sunset Crater-Cosnino analysis only.
[b] Number of clones from the indicated soil source found to cluster within the group. WI, Wisconsin pasture (4); AU, Australian forest (20, 30, 33); JA, Japanese soybean field (38); WA, Washington grassland (18).
[c] Three additional WI clones were found to group within the Y-O-G-A-*Acidobacterium* supergroup; however, their precise group placements are uncertain.
[d] Clone S124 shows low (<80%) sequence identity to other sequences of this group and may or may not constitute a fifth member.

***Acidobacterium* group.** None of the Sunset Crater-Cosnino sequences recovered in this study grouped specifically with that of the cultivated organism *A. capsulatum* (14). However, this species repeatably falls within the Y-O-G-A lineage, branching at the base of the A group. Sequences of several soil-derived clones from other studies (cluster VI MC22 [42], 103 [33], and R6-136 [4]) appear to form a cluster having *A. capsulatum* as its deepest branch and showing a specific relationship to Sunset Crater-Cosnino group A.

Analysis of full-length sequences. In order to more accurately identify the relationship between these novel groups and previously recognized bacterial divisions, more extensive phylogenetic analyses were performed on 10 nearly full-length (approximately 1,500-nt) sequences (Fig. 3). Overall, the topology of sequences of cultivated taxa is much like that previously shown for the analysis of a much larger data set (473 taxa) (26). Branching orders within and among the Y, O, G, A, and B groups match that of the tree in Fig. 2. In addition, bootstrap support is very high (99 to 100%) for monophyly of each of the Y, G, A, and B groups. The O group is represented by a single sequence, S111, but is clearly a lineage distinct from the surrounding groups. Bootstrap analyses showed low to moderate support (55 to 82%) for grouping of the A and *Acidobacterium* groups, as well as for clustering of the Y, O, and A groups to the exclusion of group G (54 to 79% support).

The large analysis of Fig. 2 suggested that the Y, O, G, A, and *Acidobacterium* clades constitute a monophyletic group of considerable phylogenetic diversity. The more extensive analyses of Fig. 3 support grouping these sequences, although bootstrap percentages for this "supergroup" are variable (maximum likelihood, 53%; maximum parsimony, 63%; neighbor joining, 96%). However, no alternative topology had more than 15% support by any analytical method. In addition, all best or most-parsimonious trees under all conditions of analysis (data not shown) display the monophyletic grouping seen in Fig. 3. This includes analyses of data sets of alternative taxon composition, with and without rate correction methods.

DISCUSSION

Our rDNA sequence-based analysis of two soils from the arid southwestern United States reveals considerable bacterial diversity that spans and greatly exceeds the known bacterial

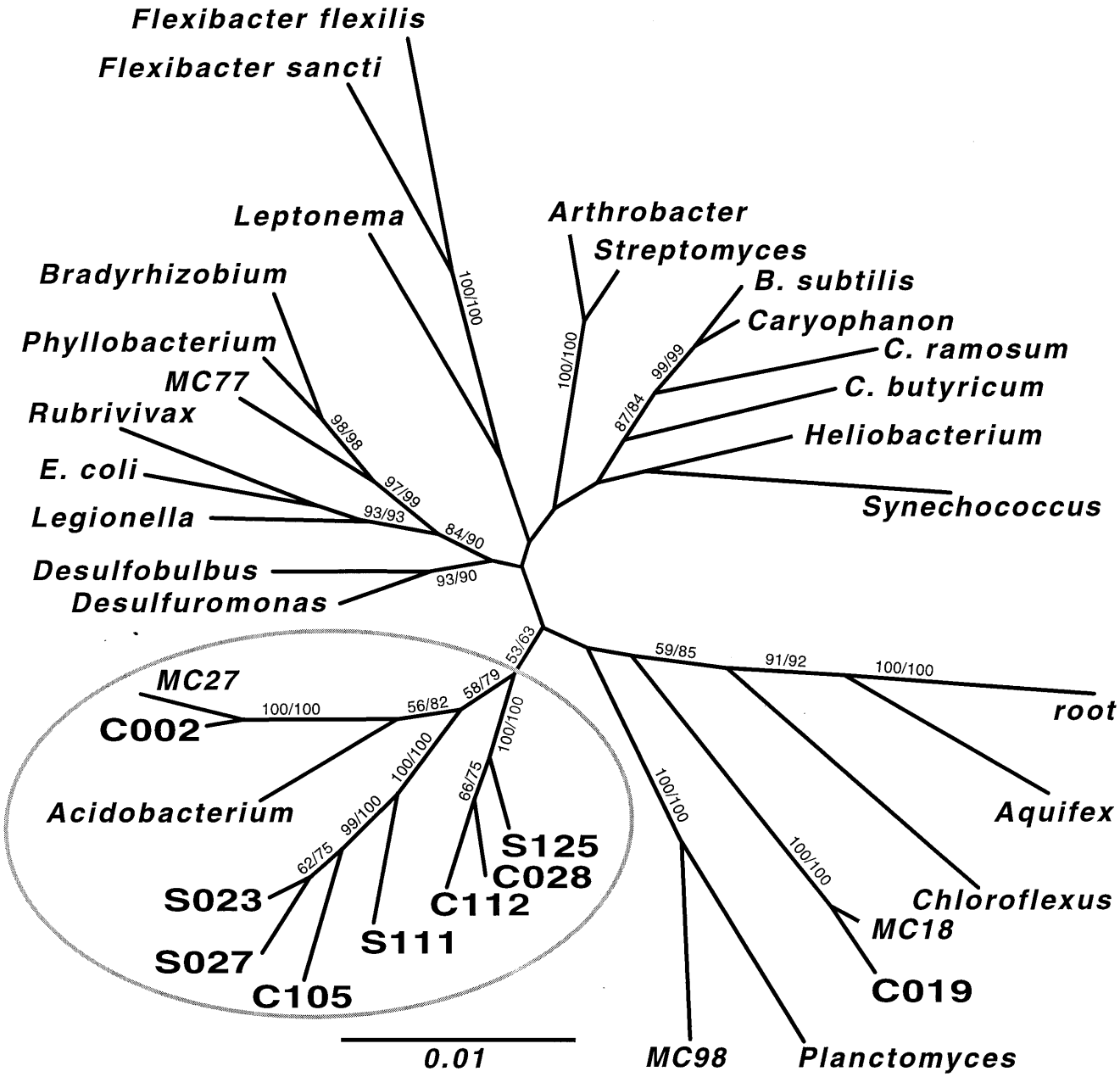

FIG. 3. Phylogenetic tree of the domain *Bacteria*, showing the positions of the Y, O, G, A, and B groups within the domain. Sequences representative of these groups are as follows: Y, S023, S027, and C105; O, S111; G, C112, C028, and S125; A, C002 and MC27; and B, C019 and MC18. The oval encircles the larger supergroup of Y, O, G, A, and *Acidobacterium* sequences, as described in the text. The tree was inferred by maximum-likelihood analysis of homologous nucleotide positions of a sequence from each organism or clone. Numbers indicate percentages of bootstrap resamplings that support branches in maximum-likelihood (before slash) and maximum-parsimony (after slash) analyses. Bootstrap results are reported only for those branches that attained >60% support with at least one of the methods used. The tree was rooted with the archaeal sequences described in the legend to Fig. 2. The bacterial species used were as described in the legend to Fig. 2, along with *Heliobacterium chlorum*, *Caryophanon latum*, and *Desulfobulbus propionicus*.

domain and thus reshapes the bacterial phylogenetic tree. Most of the soil sequences from Sunset Crater and Cosnino cluster into five groups, designated Y, O, G, A, and B. These groups are quite phylogenetically diverse and may include many novel genera of organisms covering a wide spectrum of physiological types. The preponderance of such clones in our rDNA libraries suggests that these groups of organisms may be abundant in their environments. These groups also contain sequences obtained from soils in other geographic regions of the world (4, 18, 20, 33, 38), but they do not contain any previously described cultured members.

Extensive analysis of these novel sequences has revealed a sizable addition to the bacterial phylogenetic tree in the form of a new, probably monophyletic supergroup made up of the Y, O, G, A, and *Acidobacterium* groups. Although bootstrap support for monophyly is variable, this grouping was seen in all best and most-parsimonious trees. The group branches as deeply from the main bacterial line of descent as the recog-

nized bacterial divisions, and support for the group was often better than that seen for the *Proteobacteria* and gram-positive divisions (this analysis and reference 39). The short, moderately supported branch leading to this group may be a result of the rapid radiation of bacterial lineages which obscures the branching order of many previously described divisions (39, 45). Nevertheless, support for monophyly of the individual groups is nearly absolute, indicating that they are truly distinct from previously known lineages.

None of the sequences obtained from organisms in these soils matched known sequences from any cultivated species. The similarity of these soil clones with 16S sequences from cultured organisms varied from 72.4 to 99.6% and thus spanned the range of strain-level to division-level relationships. Most of the soil sequences cluster into groups that are only distantly related to previously defined clades.

Approximately 32% of the sequences obtained from the Arizona soils group with previously recognized bacterial divisions (26). Seven of these appear to be quite closely related to sequences from the cultivable genera *Rubrivivax*, *Bradyrhizobium*, *Arthrobacter*, *Bacillus*, and *Flexibacter* and may derive from species within those genera. These genera can be cultured from many types of soils. The remainder of the Sunset Crater-Cosnino soil sequences that fall within established groups constitute distinct lineages distant from known species. These include seven sequences that follow novel lines of descent, as well as sequences in two groups previously identified from rDNA analyses of Australian forest soil: soil cluster IV, associated with the α-1 subclass of the *Proteobacteria* (33), and soil cluster II, somewhat distantly related to the planctomycetes (20). The initial finding in soil of sequences related to the planctomycetes was unexpected, but the recovery of the nine clones of cluster II from Australian soil, together with two clones from Sunset Crater, supports the suggestion that this novel group may be diverse and widespread in soils.

To date, five comparable rDNA-based analyses of bacteria from soils have been reported (4, 18, 20, 33, 38). Unlike the case for the present study, most of the sequences obtained in those studies group with previously identified clades. However, each study also recovered many novel sequences, and most (60 to 75%) of these fall into the Y, G, A, *Acidobacterium*, and B groups found in the present study. All soils so far investigated have detected members of the B group. In addition to being found in the Sunset Crater-Cosnino soils, A group organisms were detected in soils from Wisconsin, Japan, and Australia, and the G group is present in Wisconsin and Japan. The Arizona and Wisconsin soils both contain members of the Y group, while the O group has so far been detected only in the Arizona soils. Thus, these organisms may account for a significant portion of the well-documented nonculturable species predominant in soils worldwide. Preliminary analyses utilizing group-specific PCR primers (not shown) indicate that members of these groups also inhabit soils from across the United States, spanning agricultural, marine sediment, forest, and desert environments.

While there is considerable overlap among the novel sequences recovered in these soil analyses, there are striking differences between the soils studied. The soils analyzed are from vastly different geographic regions: Australia, Japan, and the temperate Midwest, the arid Northwest, and the high-desert Southwest United States. They support varied plant communities in both natural (grassland, Australian forest, and Arizona pinyon-juniper woodland) and agricultural (clover pasture and soybean field) settings. In cases where the information has been reported, the soils differ greatly in chemical and physical properties. Yet all of these soils contain members of at least some of the Y, O, G, A, and B groups. The presence of these bacterial groups in most of these soils indicates that they may be an important component of surface soils. It also suggests that individual bacterial species within these broad groups may differ in their requirements for soil chemical conditions and nutrients.

A number of potential sources of bias exist in DNA sequence-based analyses of natural communities. These have been discussed extensively elsewhere (11, 40) and include differential lysis of soil species (23), the presence of free DNA or DNA in spores (37), and preferential amplification of specific templates due to PCR primer choice (35), GC content of rDNA sequences (29), and/or rRNA gene copy number (9). The five comparable analyses of soil-derived 16S sequences reported to date (this study and references 4, 18, 20, and 38) used different DNA extraction, cloning, and PCR techniques yet recovered clones representing the same novel groups of bacteria. This suggests either that the same bias is manifest in these studies despite the use of different analytical methods or that these organisms are truly cosmopolitan and abundant in their environments. We are currently evaluating these alternatives by direct in situ probing and enumeration of soil organisms, utilizing fluorescently labeled oligonucleotide probes specific for each of the novel groups (6, 12).

We have attempted to identify cultivable species related to members of the Y, O, G, A, and B groups by screening over 600 bacterial isolates from the cinder and sandy loam soils by restriction fragment length polymorphism analysis (24) of rDNAs (data not shown), but to date we have been unsuccessful. In situ identification and cell isolation attempts are under way to begin to understand more about the ecology and physiology of these novel organisms.

ACKNOWLEDGMENTS

We gratefully acknowledge Kitty Gehring, Tom Whitham, and Paul Keim for their collaboration in studying the microbial ecology of Sunset Crater and surrounding field sites. We thank Kaysie Banton, Dante Adorada, Patricia Harper, Karen Hill, and Shannon Takala for technical assistance; John Dunbar, Jim Herrick, Karen Hill, and Norman Pace for helpful comments on the manuscript; and Mark Muntz for providing computer time.

This work was supported by a J. Robert Oppenheimer Postdoctoral Fellowship to S.M.B. and by the U.S. Department of Energy, Office of Health and Environmental Research Program for Ecosystem Research.

ADDENDUM

After this paper was submitted for review, two studies reported analyses of rDNA sequences from sediments (27a, 44). We included their data in additional phylogenetic analysis and found that Carolina bay sediments (44) contained three members of the *Acidobacterium* group, five members of the A group, and seven members of the B group. A marine sediment near a hydrocarbon seep (27a) contained four sequences that form a distinct, deeply branching clade within the Y-O-G-A-*Acidobacterium* supergroup. These additional studies expand the known geographic locations for this new supergroup to include both freshwater and marine sediments.

REFERENCES

1. **Amman, R. I., W. Ludwig, and K.-H. Schleifer.** 1995. Phylogenetic identification and in situ detection of individual microbial cells without cultivation. Microbiol. Rev. **59:**143–169.
2. **Barns, S. M., C. F. Delwiche, J. D. Palmer, and N. R. Pace.** 1996. Perspectives on archaeal diversity, thermophily and monophyly from environmental rRNA sequences. Proc. Natl. Acad. Sci. USA **93:**9188–9193.
3. **Benlloch, S., F. Rodriquez-Valera, and A. J. Martinez-Murcia.** 1995. Bacte-

rial diversity in two coastal lagoons deduced from 16S rDNA PCR amplification and partial sequencing. FEMS Microbiol. Ecol. **18:**267–280.
4. **Borneman, J., P. W. Skroch, K. M. O'Sullivan, J. A. Palus, N. G. Rumjanek, J. L. Jansen, J. Nienhuis, and E. R. Triplett.** 1996. Molecular microbial diversity of an agricultural soil in Wisconsin. Appl. Environ. Microbiol. **62:**1935–1943.
5. **Cobb, N. S., S. Mopper, C. A. Gehring, M. Caouette, K. M. Christensen, and T. G. Whitham.** 1997. Increased moth herbivory associated with environmental stress of pinyon pine at local and regional levels. Oecologia **109:**389–397.
6. **DeLong, E. F., G. S. Wickham, and N. R. Pace.** 1989. Phylogenetic stains: ribosomal RNA-based probes for the identification of single cells. Science **243:**1360–1363.
7. **Edwards, U., T. Rogall, H. Blöcker, M. Emde, and E. C. Böttger.** 1989. Isolation and direct complete determination of entire genes. Nucleic Acids Res. **17:**7843–7853.
8. **Erb, R. W., and I. Wagner-Dobler.** 1993. Detection of polychlorinated biphenyl degradation genes in polluted sediments by direct DNA extraction and polymerase chain reaction. Appl. Environ. Microbiol. **59:**4065–4073.
9. **Farrelly, V., F. A. Rainey, and E. Stackebrandt.** 1995. Effect of genome size and *rrn* gene copy number on PCR amplification of 16S rRNA genes from a mixture of bacterial species. Appl. Environ. Microbiol. **61:**2798–2801.
10. **Frothingham, R., R. L. Allen, and K. H. Wilson.** 1991. Rapid 16S ribosomal DNA sequencing from a single colony without DNA extraction or purification. BioTechniques **11:**40–44.
11. **Gray, J. P., and R. P. Herwig.** 1996. Phylogenetic analysis of the bacterial communities in marine sediments. Appl. Environ. Microbiol. **62:**4049–4059.
12. **Hahn, D., R. I. Amann, W. Ludwig, A. D. L. Akkermans, and K. Schleifer.** 1992. Detection of micro-organisms in soil using *in situ* hybridization with rRNA-targeted, fluorescently labelled oligonucleotides. J. Gen. Microbiol. **138:**879–887.
13. **Hendricks, D. M.** 1985. Arizona soils. University of Arizona Press, Tucson.
14. **Hiraishi, A., N. Kishimoto, Y. Kosako, N. Wakao, and T. Tano.** 1995. Phylogenetic position of the menaquinone-containing acidophilic chemo-organotroph *Acidobacterium capsulatum*. FEMS Microbiol. Lett. **132:**91–94.
15. **Jin, L., and M. Nei.** 1990. Limitations of the evolutionary parsimony method of phylogenetic analysis. Mol. Biol. Evol. **7:**82–103.
16. **Kopczynski, E. D., M. M. Bateson, and D. M. Ward.** 1994. Recognition of chimeric small-subunit ribosomal DNAs composed of genes from uncultivated microorganisms. Appl. Environ. Microbiol. **60:**746–748.

16a. **Kuske, C. R.** Unpublished data.

17. **Larsen, N., G. J. Olsen, B. L. Maidak, M. J. McCaughey, R. Overbeek, T. J. Macke, T. L. Marsh, and C. R. Woese.** 1993. The Ribosomal Database Project. Nucleic Acids Res. **21:**3021–3023.
18. **Lee, S.-Y., J. Bollinger, D. Bezdicek, and A. Ogram.** 1996. Estimation of the abundance of an uncultured soil bacterial strain by a competitive quantitative PCR method. Appl. Environ. Microbiol. **62:**3787–3793.
19. **Liesack, W., R. Soller, T. Stewart, H. Haas, S. Giovannoni, and E. Stackebrandt.** 1992. The influence of tachytelically (rapidly) evolving sequences on the topology of phylogenetic trees—intrafamily relationships and the phylogenetic position of Planctomycetaceae as revealed by comparative analysis of 16S ribosomal RNA sequences. Syst. Appl. Microbiol. **15:**357–362.
20. **Liesack, W., and E. Stackebrandt.** 1992. Occurrence of novel groups of the domain *Bacteria* as revealed by analysis of genetic material isolated from an Australian terrestrial environment. J. Bacteriol. **174:**5072–5078.
21. **Liesack, W., H. Weyland, and E. Stackebrandt.** 1991. Potential risks of gene amplification by PCR as determined by 16S rDNA analysis of a mixed-culture of strict barophilic bacteria. Microb. Ecol. **21:**191–198.
22. **Maidak, B. L., N. Larsen, M. J. McCaughey, R. Overbeek, G. J. Olsen, K. Fogel, J. Blandy, and C. R. Woese.** 1994. The Ribosomal Database Project. Nucleic Acids Res. **22:**3485–3487.
23. **Morè, M. I., J. B. Herrick, M. C. Silva, W. C. Ghiorse, and E. L. Madsen.** 1994. Quantitative cell lysis of indigenous microorganisms and rapid extraction of microbial DNA from sediment. Appl. Environ. Microbiol. **60:**1572–1580.
24. **Moyer, C. J., F. C. Dobbs, and D. M. Karl.** 1994. Estimation of diversity and community structure through restriction fragment length polymorphism distribution analysis of bacterial 16S rRNA genes from a microbial mat at an active, hydrothermal vent system, Loihi Seamount, Hawaii. Appl. Environ. Microbiol. **60:**871–879.
25. **Ogram, A., G. S. Sayler, and T. Barkay.** 1987. The extraction and purification of microbial DNA from sediments. J. Microbiol. Methods **7:**57–66.
26. **Olsen, G. J., C. R. Woese, and R. Overbeek.** 1994. The winds of (evolutionary) change: breathing new life into microbiology. J. Bacteriol. **176:**1–6.
27. **Olsen, G. J., H. Matsuda, R. Hagstrom, and R. Overbeek.** 1994. fastDNAml: a tool for construction of phylogenetic trees of DNA sequences using maximum likelihood. Comput. Appl. Biosci. **10:**41–48.

27a. **O'Neill, K.** Personal communication.

28. **Pedersen, K., J. Arlinger, S. Ekendahl, and L. Hallbeck.** 1996. 16S rRNA gene diversity of attached and unattached bacteria in boreholes along the access tunnel to the Aspo hard rock laboratory, Sweden. FEMS Microbiol. Ecol. **19:**249–262.
29. **Reysenbach, A., L. J. Giver, G. S. Wickham, and N. R. Pace.** 1992. Differential amplification of rRNA genes by polymerase chain reaction. Appl. Environ. Microbiol. **58:**3417–3418.
30. **Rheims, H., C. Sproer, F. A. Rainey, and E. Stackebrandt.** 1996. Molecular biological evidence for the occurrence of uncultured members of the actinomycete line of descent in different environments and geographical locations. Microbiology **142:**2863–2870.
31. **Sambrook, J., E. F. Fritsch, and T. Maniatis.** 1989. Molecular cloning: a laboratory manual, 2nd ed. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
32. **Schmidt, T. M., E. F. DeLong, and N. R. Pace.** 1991. Analysis of a marine picoplankton community by 16S rRNA gene cloning and sequencing. J. Bacteriol. **173:**4371–4378.
33. **Stackebrandt, E., W. Liesack, and B. M. Goebel.** 1993. Bacterial diversity in a soil sample from a subtropical Australian environment as determined by 16S rDNA analysis. FASEB J. **7:**232–236.
34. **Staley, J. T., and A. Konopka.** 1985. Measurement of in situ activities of non-photosynthetic microorganisms in aquatic and terrestrial habitats. Annu. Rev. Microbiol. **39:**321–346.
35. **Suzuki, M. T., and S. J. Giovannoni.** 1996. Bias caused by template annealing in the amplification of mixtures of 16S rRNA genes by PCR. Appl. Environ. Microbiol. **62:**625–630.
36. **Thompson, J. D., D. G. Higgins, and T. J. Gibson.** 1994. Clustal W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position specific gap weighting and weight matrix choice. Nucleic Acids Res. **22:**4673–4680.
37. **Tsai, Y.-L., and B. H. Olson.** 1991. Rapid method for direct extraction of DNA from soil and sediments. Appl. Environ. Microbiol. **57:**1070–1074.
38. **Ueda, T., Y. Suga, and T. Matsuguchi.** 1995. Molecular phylogenetic analysis of a soil microbial community in a soybean field. Eur. J. Soil Sci. **46:**415–421.
39. **Van de Peer, Y., J.-M. Neefs, P. De Rijk, P. De Vos, and R. De Wachter.** 1994. About the order of divergence of the major bacterial taxa during evolution. Syst. Appl. Microbiol. **17:**32–38.
40. **Ward, D. M., M. M. Bateson, R. Weller, and A. L. Ruff-Roberts.** 1992. Ribosomal RNA analysis of microorganisms as they occur in nature. Adv. Microb. Ecol. **12:**219–286.
41. **Ward, D. M., R. Weller, and M. M. Bateson.** 1990. 16S rRNA sequences reveal numerous uncultured microorganisms in a natural community. Nature **345:**63–65.
42. **Wilson, K. H., and R. B. Blitchington.** 1996. Human colonic biota studied by ribosomal DNA sequence analysis. Appl. Environ. Microbiol. **62:**2273–2278.
43. **Wilson, K. H., R. B. Blitchington, and R. C. Greene.** 1990. Amplification of bacterial 16S ribosomal DNA with polymerase chain reaction. J. Clin. Microbiol. **28:**1942–1946.
44. **Wise, M. G., J. V. McArthur, and L. J. Shimkets.** 1997. Bacterial diversity of a Carolina bay as determined by 16S rRNA gene analysis: confirmation of novel taxa. Appl. Environ. Microbiol. **63:**1505–1514.
45. **Woese, C. R.** 1987. Bacterial evolution. Microbiol. Rev. **51:**221–271.